Sharp lower bounds on the extractable randomness from non-uniform sources

Škorić, Boris; Obi, Chibuzo; Verbitskiy, Evgeny; Schoenmakers, Berry

Published in:
Information and Computation

DOI:
10.1016/j.ic.2011.06.001

Publication date:
2011

Information and Computation 209 (2011) 1184–1196

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Sharp lower bounds on the extractable randomness from non-uniform sources

Boris Škorić [a,*], Chibuzo Obi [a], Evgeny Verbitskiy [b], Berry Schoenmakers [a]

[a] Eindhoven University of Technology, The Netherlands
[b] Leiden University, The Netherlands

ARTICLE INFO

Article history:
Received 4 June 2009
Revised 21 April 2011
Available online 12 June 2011

Keywords:
Leftover Hash Lemma
Universal hash function
Randomness extraction

ABSTRACT

Extraction of uniform randomness from (noisy) non-uniform sources is an important primitive in many security applications, e.g. (pseudo-)random number generators, privacy-preserving biometrics, and key storage based on Physical Unclonable Functions. Generic extraction methods exist, using universal hash functions. There is a trade-off between the length of the extracted bit string and the uniformity of the string. In the literature there are proven lower bounds on this length as a function of the desired uniformity. The best known bound involves a quantity known as smooth min-entropy. Unfortunately, there exist at least three definitions of smooth entropy. In this paper we compare three of these definitions, and we derive improved lower bounds on the extractable randomness.
We also investigate the use of almost universal hash functions, which are slightly worse at extracting randomness than universal hash functions, but are preferable in practice because they require far less resources in devices. We show that using them has negligible effect on the extractable randomness.

1. Introduction

1.1. Randomness extraction from non-uniform noisy sources

Many cryptographic applications require input bitstrings to be uniformly distributed and exactly reproducible. Cryptographic keys, for instance, have to be uniformly random in order to prevent efficient attacks; they have to be reproducible in order to allow for decryption of encrypted data, verification of signatures, successful authentication etc. Even a single bit error in a key causes failure.

Physical sources of randomness, however, are neither uniform nor noise-free. Consider for example biometric measurements. The patterns in fingerprints and iris scans do not follow a uniform distribution, and they are never exactly re-obtained when a measurement is repeated. Measurement noise can be due to many factors, e.g. differences in lighting conditions or sensor alignment, physiological changes, difference between sensors etc. Another class of physical sources that has received a lot of attention recently are the Physical Unclonable Functions (PUFs), also known as Physical One-Way Functions, Physical Random Functions and Physically Obscured Keys. PUFs can be regarded as 'non-biological biometrics' in the sense that an identification/authentication string or secret key is derived from measuring an intrinsic property of an object. Many types of physical system for use as a PUF have been described in the literature, for instance three-dimensional multiple scattering of laser light [13], reflection of laser light from paper fibers [2], randomized dielectric properties in protective coatings on integrated circuits [21], radiofrequent responses from pieces of metal wire [5] or thin-film resonators

* Corresponding author.
E-mail address: b.skoric@tue.nl (B. Škorić).

0890-5401/$ – see front matter
doi:10.1016/j.ic.2011.06.001

[23], delay times in chip components [4] and start-up values of SRAM cells [8]. Even typed passphrases can be regarded as a non-uniform noisy source in view of possible typing mistakes.

For security and/or privacy reasons it is often necessary to apply a one-way hash function to the measurement of the biometric/PUF. Usually an attacker model is adopted where attacks by insiders have to be taken into account. Hence the storage of biometric/PUF data is assumed to be public. The hashing step hides the measurement data from the attacker. However, as the measurements are noisy, it is not possible to directly apply a hash function; a single bit error in the input causes roughly 50% of the output bits to flip. Hence, an error-correction step is required first ('information reconciliation'). This is not trivial, since the redundancy data has to be stored publicly and may reveal too much privacy-sensitive information. Similarly, if PUF data is to be used as a cryptographic key, then it should be thoroughly noise-corrected first. Here, too, it is crucial that the publicly stored redundancy data does not reveal secrets (in this case the extracted key).

After information reconciliation, the step of *privacy amplification* is applied, mapping a non-uniform random string to a shorter, almost uniform string. The requirement of uniformity is obvious in the case of key extraction. Interestingly, extracting uniform bitstrings is also desirable in biometric identification and PUF-based anti-counterfeiting, applications where the identifiers are not considered to be secret. A uniform string is the most efficient way of storing the entropy present in a measurement. Furthermore, database search speed is improved. Apart from biometrics and PUF applications, the subject of privacy amplification is highly relevant to other areas, e.g. true random number generation.

The concept of a *Fuzzy Extractor* [6,7], also known as a *helper data scheme* [12], was introduced as a primitive that achieves both information reconciliation and privacy amplification.[1] The publicly stored enrolment data (a.k.a. secure sketch or helper data) suffices to reproducibly reconstruct a string from noisy measurements, yet leaks only a negligible amount of information about the measurement or the extracted key. An overview of privacy-preserving biometrics, PUFs and fuzzy extraction is given in [22].

1.2. Lower bound on the extractable information

It is, of course, important to know which key length is safe to extract from a source. 'Safe' here means that the key's distribution is sufficiently close to uniform.

Let X denote the noisy measurement, and Y the helper data sufficient to reduce the errors to a tolerable level. The quantity $\ell^{\varepsilon}_{\mathrm{ext}}(X|Y)$ is defined as the maximum achievable key length that can be extracted from X, given that the adversary knows Y, such that the distribution of the key is ε-close to uniform. (This statement is made more precise in Section 2.) Similarly, a quantity $\ell^{\varepsilon}_{\mathrm{ext}}(X)$ is defined in the case of a noise-free source X. In [15] Renner and Wolf provided a lower bound on $\ell^{\varepsilon}_{\mathrm{ext}}(X|Y)$ in terms of the *smooth min-entropy* of X given Y. Loosely speaking, 'smoothening' of a probability distribution means probing a small neighborhood of it; the smooth entropy is the entropy of a point in this neighborhood. (This is made more precise in Section 2.)

The technique of smoothening has been used before in many papers (e.g. [9]) under a variety of names. In [1], a random variable is said to have "ϵ-HILL-type pseudoentropy at least k" if there exists a random variable X' with min-entropy at least k such that the (computational) distance between X and X' is at most ϵ. However, Renner and Wolf actually defined the smooth min-entropy of X as a uniquely determined quantity (by taking the maximum value for k), which can then be studied in its own right. The lower bound given in [15] is

$$\ell^{\varepsilon}_{\mathrm{ext}}(X|Y) \geqslant \mathsf{H}^{\rho}_{\infty}(X|Y) - \log \frac{1}{(\varepsilon-\rho)^2}. \tag{1}$$

Here $\rho \in [0,\varepsilon)$ is the 'radius' of the smoothening neighborhood, and $\mathsf{H}^{\rho}_{\infty}$ is the smooth min-entropy. This bound is sharper than previous results, which do not have the smoothening degree of freedom. It was also shown in [15] that the extractable randomness can be upper-bounded by the smooth min-entropy, i.e. (1) is optimal up to a function of ε and ρ.

Eq. (1) should be read as a maximization of the right-hand size over the smoothening parameter ρ. A nonzero ρ increases the term $\mathsf{H}^{\rho}_{\infty}(X|Y)$, but gives a penalty in the logarithmic term. The optimal ρ depends on the probability distribution of X, Y.

In the above formulation, the fuzzy extraction seems at first glance to be a matter of privacy amplification only, without the information reconciliation discussed at length in Section 1.1. However, the information reconciliation is present implicitly via the helper data Y, which is always tacitly assumed to provide sufficient error-correction redundancy. The choice of Y has a big influence on $\ell^{\varepsilon}_{\mathrm{ext}}(X|Y)$, but the formalism used in this paper never makes that explicit.

It is important to note that the bound (1) is not just of theoretical interest. It guarantees that well-known generic privacy amplification methods (universal hash functions) yield a key that is ε-close to uniformity as long as the key length does not exceed the r.h.s. of (1). This is relevant even when the distribution of X, Y is precisely known and by sheer luck happens to be such that the complete min-entropy can be extracted by a tailor-made binning scheme; in practice it may be difficult to implement the scheme on a constrained device. Hence, a sharp lower bound on $\ell^{\varepsilon}_{\mathrm{ext}}$ has practical implications.

While the result (1) is very useful, some unsatisfactory issues remain. Unfortunately, at least three different definitions of smooth entropy appear in the literature [15,16,10,14,11], and it is not immediately clear if they are equivalent. While it

[1] In this setting there is no additional source of randomness to aid the privacy amplification.

seems that they can be used interchangeably as far as chain rules and similar inequalities are concerned, the consequences for bounds such as (1) are less evident.

Furthermore, the definition of smoothening employed in [15] involves distributions that are not normalized. While that is not a problem in itself, it can lead to technical difficulties when smooth entropies are applied in a context that requires normalization (see Sections 2 and 3.2).

1.3. Contributions and outline

In this paper we address the issue of the multiple definitions of smooth entropy, and we provide lower bounds on extractable randomness that are tighter than previously stated in the literature.

We introduce notation and give a list of definitions in Section 2.1. We then present a number of useful lemmas in Section 2.2. Two of these are novel, concerning the leftover hash lemma in the case of un-normalized distributions.

Our main results are stated in Section 3. We first prove a sharper lower bound on the extractable randomness in the unconditional case (Theorem 1). The improvement has two causes:

(a) We make use of the leftover hash lemma for un-normalized distributions. This has the effect of reducing the penalty term $\log(\varepsilon-\rho)^2$ to $\log\varepsilon(\varepsilon-\rho)$.
(b) We use Rényi entropy of order 2 (H_2) instead of min-entropy. In this way the bound is written in terms of the quantity that naturally appears in the leftover hash lemma, without losses from further inequalities.[2]

We illustrate the improvement in graphs. For low-entropy sources, the relative improvement in the amount of extracted random bits is significant.

We prove a bound (Theorem 2) for the conditional case that is sharper than (1). It is formulated in terms of the *smooth average conditional Rényi entropy of order two*. The average conditional Rényi entropy ($\widetilde{\mathsf{H}}_2$) is the natural quantity appearing in the leftover hash lemma.[3] We show that the above-mentioned beneficial effect (a) does not occur in the conditional case.

We compare three of the existing definitions of smooth entropy regarding their impact on provable bounds on randomness extraction. It turns out that they are not all equivalent in this respect. The two definitions with un-normalized distributions give identical results. However, the 'normalized' version of smoothening yields less sharp bounds. The underlying reason is that the normalization constraint effectively doubles the statistical distance between a distribution and its smoothened version, as compared to un-normalized smoothening.

The results above are based on the existence of *universal* families of hash functions (also called 'two-universal'). In Section 4 we study the case of *almost universal* families of hash functions [18,19]. These do not achieve the same amount of privacy amplification, but require significantly less nonvolatile storage to implement and are therefore often preferable in practice. We show that their use has a negligible effect on the extractable randomness.

2. Preliminaries

In Section 2.1 we first introduce the notation that will be used in the rest of the paper. Then we list a number of definitions in order to specify precisely which notions of distance, entropies, smoothening and extractability we are talking about.

Throughout this paper we make an effort to use notation which explicitly shows which variables are operated on by a mapping. In particular, the notation always makes it evident that an entropy is a function of a distribution (as opposed to a random variable). For instance, for a random variable X with distribution $\mathbb{P}$, we always write $\mathsf{H}(\mathbb{P})$ for the entropy instead of the more usual notation $\mathsf{H}(X)$. This will sometimes lead to awkward superscripts, but it is a price worth paying for the improved clarity. Clarity is especially needed when smooth entropies are handled. A smoothened distribution is not necessarily normalized; this makes it difficult to write statements like "variable X has distribution $\mathbb{Q}$". In order to avoid such awkwardness we concentrate on statements about distributions. We formulate all the definitions in such a way that they apply to normalized as well as un-normalized distributions.

In Section 2.2 we list some lemmas that are necessary to prove our results.

2.1. Notation and definitions

We denote stochastic variables by capitals, specific values by lowercase letters, and spaces by calligraphic symbols. Distributions are written in mathbold font. E.g. $X\in\mathcal{X}$, $X\sim\mathbb{P}$, $\Pr[X=x]=\mathbb{P}(x)$. The notation $\mathcal{P}_{\mathcal{X}}$ denotes the space of non-negative functions on $\mathcal{X}$. As a distance measure between distributions we use the usual definition of statistical distance, generalized to arbitrary non-negative functions. The notation $U_{\mathcal{X}}$ stands for the uniform distribution over a set $\mathcal{X}$.

[2] The use of H_2 is not new in itself, nor is the smooth variant H_2^ρ, and replacing H_∞ by H_2^ρ is straightforward. However, we want to stress that the results in the literature, which are usually stated in terms of min-entropy, can be improved upon when every last bit counts.

[3] It was already pointed out in [7] that worst-case conditioning in the conditional min-entropy $H_\infty(X|Y)$ can be replaced by average-case conditioning.

Definition 1 *(Statistical distance).* Let $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_{\mathcal{X}}$. The statistical distance between $\mathbb{P}$ and $\mathbb{Q}$ is defined as

$$\Delta(\mathbb{P}, \mathbb{Q}) = \frac{1}{2} \sum_{x \in \mathcal{X}} \big| \mathbb{P}(x) - \mathbb{Q}(x) \big|.$$

For $(X, Y) \sim \mathbb{P}$, the marginal distribution of X is denoted as $\mathbb{P}_1$, with $\Pr[X = x] = \mathbb{P}_1(x) = \sum_y \mathbb{P}(x, y)$. Similarly, $\Pr[Y = y] = \mathbb{P}_2(y) = \sum_x \mathbb{P}(x, y)$. The conditional probability $\Pr[X = x \mid Y = y]$ is denoted as $\mathbb{P}_{1|2}(x, y) = \mathbb{P}(x, y)/\mathbb{P}_2(y)$.[4] The advantage of this notation is that marginals and conditioning can be easily written down for un-normalized distributions.

Definition 2 *(Rényi entropy).* Let $\mathbb{P} \in \mathcal{P}_{\mathcal{X}}$. For $\alpha \in (0, 1) \cup (1, \infty)$, the Rényi entropy of $\mathbb{P}$ is defined as

$$\mathsf{H}_\alpha(\mathbb{P}) = -\frac{1}{\alpha - 1} \log \sum_{x \in \mathcal{X}} \big[\mathbb{P}(x)\big]^\alpha.$$

For $0 \leqslant \alpha \leqslant \beta$ it holds that $\mathsf{H}_\beta(\mathbb{P}) \leqslant \mathsf{H}_\alpha(\mathbb{P})$.

Definition 3 *(Min-entropy).* Let $\mathbb{P} \in \mathcal{P}_{\mathcal{X}}$. The min-entropy of $\mathbb{P}$ is defined as

$$\mathsf{H}_\infty(\mathbb{P}) = -\log \max_{x \in \mathcal{X}} \mathbb{P}(x).$$

Definition 4 *(Conditional Rényi entropy).* Let $\mathbb{P} \in \mathcal{P}_{\mathcal{X} \times \mathcal{Y}}$. Let $\alpha \in (0, 1) \cup (1, \infty)$. The conditional Rényi entropy is defined as

$$\mathsf{H}_{\alpha(1|2)}(\mathbb{P}) = -\frac{1}{\alpha - 1} \log \max_{y \in \operatorname{supp} \mathbb{P}_2} \sum_{x \in \mathcal{X}} \big[\mathbb{P}_{1|2}(x, y)\big]^\alpha.$$

Definition 5 *(Conditional min-entropy).* Let $\mathbb{P} \in \mathcal{P}_{\mathcal{X} \times \mathcal{Y}}$. The conditional min-entropy of $\mathbb{P}$ is defined as

$$\mathsf{H}_{\infty(1|2)}(\mathbb{P}) = -\log \max_{x \in \mathcal{X}} \max_{y \in \operatorname{supp} \mathbb{P}_2} \mathbb{P}_{1|2}(x, y).$$

For $(X, Y) \sim \mathbb{P}$ the conditional min-entropy is usually denoted as $H_\infty(X|Y)$.

Definition 6 *(Average conditional min-entropy).* (From [7].) Let $\mathbb{P} \in \mathcal{P}_{\mathcal{X} \times \mathcal{Y}}$. The average conditional min-entropy of $\mathbb{P}$, with conditioning on the second argument, is defined as

$$\widetilde{\mathsf{H}}_{\infty(1|2)}(\mathbb{P}) = -\log \sum_{y \in \mathcal{Y}} \mathbb{P}_2(y) \max_{x \in \mathcal{X}} \mathbb{P}_{1|2}(x, y) = -\log \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} \mathbb{P}(x, y).$$

The usual notation is $\widetilde{\mathsf{H}}_\infty(X|Y)$ for $(X, Y) \sim \mathbb{P}$.

Definition 7 *(Average conditional Rényi entropy).* Let $\mathbb{P}$ be a probability measure on $\mathcal{X} \times \mathcal{Y}$, and let $(X, Y) \sim \mathbb{P}$. Let $\alpha \in (0, 1) \cup (1, \infty)$. The average conditional Rényi entropy of X given Y is defined as

$$\widetilde{\mathsf{H}}_{\alpha(1|2)}(\mathbb{P}) = -\frac{1}{\alpha - 1} \log \sum_{y \in \mathcal{Y}} \mathbb{P}_2(y) \sum_{x \in \mathcal{X}} \big[\mathbb{P}_{1|2}(x, y)\big]^\alpha.$$

The following two definitions are somewhat awkward, but they allow us to generalize stochastic variables like $f(X)$ to the case of un-normalized distributions.

Definition 8 *(Induced action on a distribution).* Let $f : \mathcal{R} \times \mathcal{X} \to \mathcal{Z}$ be a function, with the first argument drawn uniformly random from $\mathcal{R}$. Let $\mathbb{P} \in \mathcal{P}_{\mathcal{X}}$ and $X \sim \mathbb{P}$. We define the induced action of f on $\mathbb{P}$ as a mapping $f^* : \mathcal{P}_{\mathcal{X}} \to \mathcal{P}_{\mathcal{R} \times \mathcal{Z}}$, with

$$\big(f^* \mathbb{P}\big)(r, t) = \Pr\big[R = r,\ f(R, X) = t\big] = \frac{1}{|\mathcal{R}|} \sum_{x \in \mathcal{X}:\ f(r,x) = t} \mathbb{P}(x).$$

[4] We use the convention $\mathbb{P}_{1|2}(x, y) = 0$ when $\mathbb{P}_2(y) = 0$.

Definition 9 *(Induced action on a joint distribution).* Let $F : \mathcal{R} \times \mathcal{X} \to \mathcal{Z}$ be a function, with the first argument drawn uniformly random from $\mathcal{R}$. Let $\mathbb{P} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}$ and $(X, Y) \sim \mathbb{P}$. We define the induced action of F on $\mathbb{P}$ as a mapping $F^* : \mathcal{P}_{\mathcal{X}\times\mathcal{Y}} \to \mathcal{P}_{\mathcal{Y}\times\mathcal{R}\times\mathcal{Z}}$, with

$$\big(F^*\mathbb{P}\big)(y, r, t) = \Pr\big[Y = y,\ R = r,\ F(R, X) = t\big] = \frac{1}{|\mathcal{R}|}\mathbb{P}_2(y) \sum_{x\in\mathcal{X}:\ F(r,x)=t} \mathbb{P}_{1|2}(x, y).$$

Next we show three different notions of smooth entropy that appear in the literature. They differ in the sort of modifications to the probability distribution that are allowed. In order to distinguish between them we introduce the terms 'strictly bounded', 'loosely bounded' and 'normalized'.

Definition 10 *(Strictly bounded vicinity of a distribution).* (From [11].) Let $\mathbb{P}$ be a probability measure on $\mathcal{X}$. Let $\rho \geqslant 0$. We define the strictly bounded ρ-vicinity of $\mathbb{P}$ as

$$B^{\rho}_{\text{strict}}(\mathbb{P}) = \left\{\mathbb{Q} \in \mathcal{P}_{\mathcal{X}}\colon\ \forall_{x\in\mathcal{X}}\mathbb{Q}(x) \leqslant \mathbb{P}(x),\ \sum_{x\in\mathcal{X}} \mathbb{Q}(x) \geqslant 1 - \rho\right\}.$$

Note that $\mathbb{Q} \in B^{\rho}(\mathbb{P})$ with $\mathbb{Q} \neq \mathbb{P}$ is not a probability measure.

Definition 11 *(Normalized vicinity of a distribution).* Let $\mathbb{P}$ be a probability measure on $\mathcal{X}$. Let $\rho \geqslant 0$. The normalized ρ-vicinity of $\mathbb{P}$ is defined as

$$B^{\rho}_{\text{norm}}(\mathbb{P}) = \left\{\mathbb{Q} \in \mathcal{P}_{\mathcal{X}}\colon\ \Delta(\mathbb{Q}, \mathbb{P}) \leqslant \rho,\ \sum_{x\in\mathcal{X}} \mathbb{Q}(x) = 1\right\}.$$

Definition 12 *(Loosely bounded vicinity of a distribution).* Let $\mathbb{P}$ be a probability measure on $\mathcal{X}$. Let $\rho \geqslant 0$. The loosely bounded ρ-vicinity of $\mathbb{P}$ is defined as

$$B^{\rho}_{\text{loose}}(\mathbb{P}) = \left\{\mathbb{Q} \in \mathcal{P}_{\mathcal{X}}\colon\ \Delta(\mathbb{Q}, \mathbb{P}) \leqslant \frac{\rho}{2},\ \sum_{x\in\mathcal{X}} \mathbb{Q}(x) \leqslant 1\right\}.$$

The strictly bounded version appeared in [15] (unnumbered definition) and [11]. The loosely bounded version appeared in [14, Definition 3.2.1]. The normalized version was used in [10]. (Yet another definition features in [16], but we will not discuss it here.)

For all three types of vicinity, the smooth entropy is defined analogously. In the definitions below, the subscript 'type' denotes the type of vicinity. The result (1) from [15] was formulated in terms of the 'strictly bounded' vicinity.

Definition 13 *(Smooth Rényi entropy).* Let $\mathbb{P}$ be a probability measure on $\mathcal{X}$. Let $\rho \geqslant 0$. The smooth Rényi entropy of $\mathbb{P}$ is defined as

$$\mathsf{H}^{\rho,\text{type}}_{\alpha}(\mathbb{P}) = \max_{\mathbb{Q}\in B^{\rho}_{\text{type}}(\mathbb{P})} \mathsf{H}_{\alpha}(\mathbb{Q}).$$

Definition 14 *(Smooth conditional Rényi entropy).* Let $\mathbb{P}$ be a probability measure on $\mathcal{X} \times \mathcal{Y}$. Let $\rho \geqslant 0$ and $\alpha \in (0, 1) \cup (1, \infty)$. The ρ-smooth conditional Rényi entropy of order α of $\mathbb{P}$ is defined as

$$\mathsf{H}^{\rho,\text{type}}_{\alpha(1|2)}(\mathbb{P}) = \max_{\mathbb{Q}\in B^{\rho}_{\text{type}}(\mathbb{P})} \frac{1}{1-\alpha} \log \max_{y\in\operatorname{supp}\mathbb{P}_2} \sum_{x\in\mathcal{X}} \left[\frac{\mathbb{Q}(x, y)}{\mathbb{P}_2(y)}\right]^{\alpha}.$$

Definition 15 *(Smooth conditional min-entropy).* Let $\mathbb{P}$ be a probability measure on $\mathcal{X} \times \mathcal{Y}$. Let $\rho \geqslant 0$. The smooth conditional min-entropy is defined as

$$\mathsf{H}^{\rho,\text{type}}_{\infty(1|2)}(\mathbb{P}) = \max_{\mathbb{Q}\in B^{\rho}_{\text{type}}(\mathbb{P})} -\log \max_{y\in\operatorname{supp}\mathbb{P}_2} \max_{x\in\mathcal{X}} \frac{\mathbb{Q}(x, y)}{\mathbb{P}_2(y)}.$$

The usual notation is $\mathsf{H}^{\rho}_{\infty}(X|Y)$ for $(X, Y) \sim \mathbb{P}$. Note that the smooth entropies in Definitions 14 and 15 cannot be written as a maximization $\max_{\mathbb{Q}}$ of some expression in $\mathbb{Q}$ only. It will turn out that we need such a property later on; hence we introduce the following definition.

Definition 16 *(Smooth average conditional Rényi entropy).* Let $\mathbb{P}$ be a probability measure on $\mathcal{X} \times \mathcal{Y}$. Let $\rho \geqslant 0$ and $\alpha \in (0,1) \cup (1,\infty)$. The ρ-smooth average conditional Rényi entropy of order α of $\mathbb{P}$ is defined as

$$\widetilde{\mathrm{H}}^{\rho,\mathrm{type}}_{\alpha(1|2)}(\mathbb{P}) = \max_{\mathbb{Q}\in B^{\rho}_{\mathrm{type}}(\mathbb{P})} \widetilde{\mathrm{H}}_{\alpha(1|2)}(\mathbb{Q}).$$

The smooth average conditional min-entropy follows by setting $\alpha = \infty$. Next we give the formal definition of extractable randomness. This quantity plays a central role in Section 3.

Definition 17 *(Extractable randomness).* (See Definition 3 in [15].) Let $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ be random variables, with $(X,Y) \sim \mathbb{P}$. Let R be a random variable independent of X and Y, uniformly distributed on some set $\mathcal{R}$. For any $\varepsilon > 0$ we say that a finite set $\mathcal{Z}$ is ε-allowed if there exists a function $F : \mathcal{X} \times \mathcal{R} \to \mathcal{Z}$ such that

$$\Delta\big(F^*\mathbb{P}, \mathbb{P}_2 \times U_{\mathcal{R}} \times U_{\mathcal{Z}}\big) \leqslant \varepsilon.$$

The ε-extractable randomness of X conditioned on Y is defined as

$$\ell^{\varepsilon}_{\mathrm{ext}}(X|Y) = \max\big\{\log|\mathcal{Z}|\colon\ \mathcal{Z} \text{ is } \varepsilon\text{-allowed}\big\}.$$

The definition of extractable randomness in the unconditional case follows from Definition 17 by omitting Y. Lower bounds on the extractable randomness are proven using *universal hash functions*.

Definition 18 *(Universal family of hash functions).* (See [3].) Let $\mathcal{R}$, $\mathcal{X}$ and $\mathcal{T}$ be finite sets. Let $\{\Phi_r\}_{r\in\mathcal{R}}$ be a family of hash functions from $\mathcal{X}$ to $\mathcal{T}$. The family $\{\Phi_r\}_{r\in\mathcal{R}}$ is called universal iff, for R drawn uniformly from $\mathcal{R}$, it holds that

$$\Pr\big[\Phi_R(x) = \Phi_R\big(x'\big)\big] \leqslant 1/|\mathcal{T}|$$

for all $x, x' \in \mathcal{X}$ with $x' \neq x$.

Such a family is alternatively known as 'two-universal'. We also give the definition of a more general class of hash functions.

Definition 19 *(Almost universal family of hash functions).* (See [18,19].) Let $\eta \geqslant 0$. Let $\mathcal{R}$, $\mathcal{X}$ and $\mathcal{T}$ be finite sets. Let $\{\Phi_r\}_{r\in\mathcal{R}}$ be a family of hash functions from $\mathcal{X}$ to $\mathcal{T}$. The family $\{\Phi_r\}_{r\in\mathcal{R}}$ is called η-almost universal iff, for R drawn uniformly from $\mathcal{R}$, it holds that

$$\Pr\big[\Phi_R(x) = \Phi_R\big(x'\big)\big] \leqslant \eta$$

for all $x, x' \in \mathcal{X}$ with $x' \neq x$.

Note that a $1/|\mathcal{T}|$-almost universal family of hash functions is universal. The above defined classes of hash functions play a central role in this paper: They are used in Lemmas 4 and 5 in Section 2.2, on which our main results are based.

2.2. Lemmas

Lemma 1 *(Jensen's inequality for concave functions). Let φ be a real concave function. Let n be a positive integer. Let $a_1, \ldots, a_n$ be positive weights and $x_1, \ldots, x_n$ be real numbers. Then*

$$\varphi\left(\frac{\sum_{i=1}^n a_i x_i}{\sum_{i=1}^n a_i}\right) \geqslant \frac{\sum_{i=1}^n a_i \varphi(x_i)}{\sum_{i=1}^n a_i}.$$

Lemma 2. *(See Exercise 8.36 in [17].) Let $q_1, \ldots, q_m$ be real numbers satisfying $\sum_{s=1}^m q_s = 1$. Then it holds that*

$$\sum_{s=1}^m q_s^2 \geqslant \frac{1}{m}.$$

Proof. $0 \leqslant \sum_s (q_s - \frac{1}{m})^2 = \sum_s q_s^2 - \frac{2}{m}\sum_s q_s + \sum_s 1/m^2 = \sum_s q_s^2 - \frac{1}{m}$. □

Lemma 3. *(Generalization of Theorem 8.36 in [17] to un-normalized distributions.) Let $\mathbb{Q} \in \mathcal{P}_{\mathcal{Z}}$ such that $\sum_{z\in\mathcal{Z}} \mathbb{Q}(z) = 1-\alpha$. Then the statistical distance between $\mathbb{Q}$ and the uniform distribution U can be bounded as*

$$\Delta(\mathbb{Q}, U) \leqslant \frac{1}{2}\sqrt{|\mathcal{Z}| \sum_{z\in\mathcal{Z}} \mathbb{Q}^2(z) - 1 + 2\alpha}. \tag{2}$$

Proof. Define $D = \Delta(\mathbb{Q}, U)$ and $q_z := |\mathbb{Q}(z) - \frac{1}{|\mathcal{Z}|}|/2D$ satisfying $\sum_z q_z = 1$. Application of Lemma 2 yields $\frac{1}{|\mathcal{Z}|} \leqslant \frac{1}{4D^2}[\sum_z \mathbb{Q}^2(z) - \frac{2}{|\mathcal{Z}|}\sum_z \mathbb{Q}(z) + \sum_z |\mathcal{Z}|^{-2}] = \frac{1}{4D^2}[\sum_z \mathbb{Q}^2(z) - \frac{2}{|\mathcal{Z}|}(1-\alpha) + \frac{1}{|\mathcal{Z}|}]$. Rearranging the inequality gives (2). □

Lemma 4 *(Leftover hash lemma in the unconditional case). (Generalization of Theorem 8.37 in [17] to un-normalized distributions.) Let $\alpha \in [0,1)$. Let $\mathcal{X}$ be a finite set. Let $\mathbb{Q} \in \mathcal{P}_{\mathcal{X}}$ satisfy $\sum_{x\in\mathcal{X}} \mathbb{Q}(x) = 1-\alpha$. Let $f : \mathcal{R}\times\mathcal{X} \to \{0,1\}^\ell$ be an η-almost universal hash function (see Definition 19), with $\eta = (1+\delta)2^{-\ell}$, and with r drawn uniformly random from $\mathcal{R}$. Let $f^*\mathbb{Q}$ be the induced action of f on $\mathbb{Q}$, as given in Definition 8. Then*

$$\Delta\big(f^*\mathbb{Q}, U_{\mathcal{R}} \otimes U_\ell\big) \leqslant \frac{1}{2}\sqrt{\alpha^2 + \delta + 2^{\ell - \mathsf{H}_2(\mathbb{Q})}}, \tag{3}$$

where $U_{\mathcal{R}}$ stands for the uniform distribution on $\mathcal{R}$ and U_ℓ stands for the uniform distribution on $\{0,1\}^\ell$.

A proof is given in Appendix A.

Lemma 5 *(Leftover hash lemma in the conditional case). (Generalization of Exercise 8.71 in [17] to un-normalized distributions.) Let $\beta \in [0,1)$. Let $\mathbb{Q} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}$ such that $\sum_{x\in\mathcal{X}, y\in\mathcal{Y}} \mathbb{Q}(x,y) = 1-\beta$. Let $F : \mathcal{R}\times\mathcal{X} \to \{0,1\}^\ell$ be an η-almost universal hash function, with $\eta = (1+\delta)2^{-\ell}$, and with r drawn randomly from $\mathcal{R}$. Let $F^*\mathbb{Q}$ be the induced action of F on $\mathbb{Q}$ as given in Definition 9. Then*

$$\Delta\big(F^*\mathbb{Q}, \mathbb{Q}_2 \otimes U_{\mathcal{R}} \otimes U_\ell\big) \leqslant \frac{1}{2}\sqrt{1-\beta}\sqrt{\delta + 2^{\ell - \widetilde{\mathsf{H}}_{2(1|2)}(\mathbb{Q})}}.$$

We provide a proof in Appendix A. Remark: Lemmas 4 and 5 can also be formulated with H_∞ instead of H_2, since $\mathsf{H}_2(\mathbb{Q}) \geqslant \mathsf{H}_\infty(\mathbb{Q})$.

Lemma 6. *Let $\mathbb{P}$ be a probability measure on $\mathcal{X}$. Let $\rho \geqslant 0$. Then for any $\mathbb{Q} \in B^\rho_{\text{strict}}(\mathbb{P})$*

$$\Delta(\mathbb{P}, \mathbb{Q}) \leqslant \frac{1}{2}\rho.$$

Proof. $\Delta(\mathbb{P}, \mathbb{Q}) = \frac{1}{2}\sum_{x\in\mathcal{X}} |\mathbb{P}(x) - \mathbb{Q}(x)| = \frac{1}{2}[1 - \sum_{x\in\mathcal{X}} \mathbb{Q}(x)] \leqslant \frac{1}{2}\rho$. □

Lemma 7. *Let $\mathbb{P}$ be a probability measure on $\mathcal{X}\times\mathcal{Y}$. Let $\rho \geqslant 0$. Then for any $\mathbb{Q} \in B^\rho_{\text{strict}}(\mathbb{P})$*

$$\Delta(\mathbb{P}_2, \mathbb{Q}_2) \leqslant \frac{1}{2}\rho.$$

Proof. $\Delta(\mathbb{P}_2, \mathbb{Q}_2) = \frac{1}{2}\sum_y |\sum_x \mathbb{P}(x,y) - \sum_x \mathbb{Q}(x,y)| = \frac{1}{2}[1 - \sum_{xy} \mathbb{Q}(x,y)] \leqslant \frac{1}{2}\rho$. □

Lemma 8. *Let $\mathbb{P}$ be a probability measure on $\mathcal{X}\times\mathcal{Y}$. Let $\rho \geqslant 0$. Then it holds that*

$$\widetilde{\mathsf{H}}^{\rho,\text{type}}_{2(1|2)}(\mathbb{P}) \geqslant \mathsf{H}^{\rho,\text{type}}_{\infty(1|2)}(\mathbb{P}).$$

Proof. We use that $\widetilde{\mathsf{H}}^{\rho,\text{type}}_{2(1|2)}(\mathbb{P}) \geqslant \widetilde{\mathsf{H}}^{\rho,\text{type}}_{\infty(1|2)}(\mathbb{P})$. Moreover,

$$\begin{aligned}
\widetilde{\mathsf{H}}^{\rho,\text{type}}_{\infty(1|2)}(\mathbb{P}) &= \max_{\mathbb{Q}\in B^\rho_{\text{type}}(\mathbb{P})} \widetilde{\mathsf{H}}_{\infty(1|2)}(\mathbb{Q}) \\
&= \max_{\mathbb{Q}\in B^\rho_{\text{type}}(\mathbb{P})} -\log \sum_{y\in\mathcal{Y}} \mathbb{Q}_2(y) \max_{x\in\mathcal{X}} \mathbb{Q}_{1|2}(x,y) \\
&= \max_{\mathbb{Q}\in B^\rho_{\text{type}}(\mathbb{P})} -\log \sum_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} \mathbb{Q}(x,y)
\end{aligned}$$

$$= \max_{\mathbb{Q}\in B^{\rho}_{\text{type}}(\mathbb{P})} -\log \sum_{y\in\operatorname{supp}\mathbb{P}_2} \mathbb{P}_2(y) \max_{x\in\mathcal{X}} \frac{\mathbb{Q}(x,y)}{\mathbb{P}_2(y)}$$
$$\geqslant \max_{\mathbb{Q}\in B^{\rho}_{\text{type}}(\mathbb{P})} -\log \max_{y\in\operatorname{supp}\mathbb{P}_2} \max_{x\in\mathcal{X}} \frac{\mathbb{Q}(x,y)}{\mathbb{P}_2(y)}$$
$$= \mathsf{H}^{\rho,\text{type}}_{\infty(1|2)}(\mathbb{P}). \qquad \square$$

3. Main result

We prove two theorems regarding the lower bound on extractable randomness (see Definition 17). Both provide an improved bound in terms of the 'strictly bounded' type of smooth entropy. The first theorem addresses the unconditional case. The second theorem addresses the conditional case. The proofs are based on the existence of *universal* families of hash functions (see Definition 18). The more general (and more easy to implement) case of *almost universal* hash functions is treated in Section 4.

Theorem 1. *Let $X\sim\mathbb{P}$ be a random variable. Let $\varepsilon>0$. Then*

$$\ell^{\varepsilon}_{\text{ext}}(X) \geqslant \max_{\rho\in[0,\varepsilon)} \left[\mathsf{H}^{\rho,\text{strict}}_2(\mathbb{P}) + 2 - \log\frac{1}{\varepsilon(\varepsilon-\rho)} \right].$$

Theorem 2. *Let $\mathbb{P}$ be a probability measure on $\mathcal{X}\times\mathcal{Y}$, and $(X,Y)\sim\mathbb{P}$. Let $\varepsilon>0$. Then*

$$\ell^{\varepsilon}_{\text{ext}}(X|Y) \geqslant \max_{\rho\in[0,\varepsilon)} \left[\widetilde{\mathsf{H}}^{\rho,\text{strict}}_{2(1|2)}(\mathbb{P}) + 2 - \log\frac{1-\rho}{(\varepsilon-\rho)^2} \right].$$

Proof of Theorem 1. Consider $\hat{\mathbb{Q}}\in B^{\rho}_{\text{strict}}(\mathbb{P})$ achieving $\mathsf{H}_2(\hat{\mathbb{Q}}) = \mathsf{H}^{\rho,\text{strict}}_2(\mathbb{P})$. We explicitly use the fact that $\sum_{x\in\mathcal{X}}\hat{\mathbb{Q}}(x) = 1-\rho$. Consider a function $f:\mathcal{X}\times\mathcal{R}\to\{0,1\}^{\ell}$, with $\ell = \mathsf{H}^{\rho,\text{strict}}_2(\mathbb{P}) + 2 - \log\frac{1}{\varepsilon(\varepsilon-\rho)}$ whose output has distance from uniformity $\varepsilon-\rho/2$. Its existence is guaranteed by Lemma 4 (taking $\delta=0$ since we consider universal hash functions, and substituting $\alpha=\rho$). It is readily verified that substitution of ℓ into (3) yields statistical distance $\varepsilon-\rho/2$ from uniformity.

When f is applied to $\hat{\mathbb{Q}}$, the result is a string of length ℓ, with uniformity $\varepsilon-\rho/2$. What happens when f is applied to $\mathbb{P}$? Then the uniformity is bounded as follows

$$\begin{aligned}\Delta\big(f^*\mathbb{P}, U_{\mathcal{R}}\otimes U_{\ell}\big) &\leqslant \Delta\big(f^*\mathbb{P}, f^*\hat{\mathbb{Q}}\big) + \Delta\big(f^*\hat{\mathbb{Q}}, U_{\mathcal{R}}\otimes U_{\ell}\big)\\ &\leqslant \Delta(U_{\mathcal{R}}\otimes\mathbb{P}, U_{\mathcal{R}}\otimes\hat{\mathbb{Q}}) + (\varepsilon-\rho/2)\\ &\leqslant \rho/2 + (\varepsilon-\rho/2) = \varepsilon. \end{aligned} \tag{4}$$

In the first line we used the triangle inequality. In the second line we used the uniformity property of the extractor f. Finally in the third line we applied Lemma 6. $\square$

Proof of Theorem 2. Consider $\hat{\mathbb{Q}}\in B^{\rho}_{\text{strict}}(\mathbb{P})$ achieving $\widetilde{\mathsf{H}}_{2(1|2)}(\hat{\mathbb{Q}}) = \widetilde{\mathsf{H}}^{\rho,\text{strict}}_{2(1|2)}(\mathbb{P})$. The existence of such a $\hat{\mathbb{Q}}$ follows directly from Definition 16. Next we explicitly use that $\sum_{x\in\mathcal{X},\,y\in\mathcal{Y}}\hat{\mathbb{Q}}(x,y) = 1-\rho$. Consider a function $F:\mathcal{X}\times\mathcal{R}\to\{0,1\}^{\ell}$, acting on the source $\hat{\mathbb{Q}}$, extracting a string of length $\ell = \widetilde{\mathsf{H}}^{\rho,\text{strict}}_{2(1|2)}(\mathbb{P}) + 2 - 2\log\frac{\sqrt{1-\rho}}{\varepsilon-\rho}$ that has uniformity $\varepsilon-\rho$. The existence of such F is guaranteed by Lemma 5 (again setting $\delta=0$, and substituting $\beta=\rho$). What happens when F is applied to $\mathbb{P}$ instead of $\hat{\mathbb{Q}}$? We first apply the triangle inequality,

$$\Delta\big(F^*\mathbb{P}, \mathbb{P}_2\otimes U_{\mathcal{R}}\otimes U_{\ell}\big) \leqslant \Delta\big(F^*\mathbb{P}, F^*\hat{\mathbb{Q}}\big) + \Delta\big(F^*\hat{\mathbb{Q}}, \mathbb{P}_2\otimes U_{\mathcal{R}}\otimes U_{\ell}\big). \tag{5}$$

The first term in (5) is bounded by using the fact that conditioning reduces the statistical distance,

$$\Delta\big(F^*\mathbb{P}, F^*\hat{\mathbb{Q}}\big) \leqslant \Delta(U_{\mathcal{R}}\otimes\mathbb{P}, U_{\mathcal{R}}\otimes\hat{\mathbb{Q}}) = \Delta(\mathbb{P},\hat{\mathbb{Q}}) \leqslant \frac{1}{2}\rho. \tag{6}$$

The last term in (5) is bounded by using one more triangle inequality.

$$\begin{aligned}&\Delta\big(F^*\hat{\mathbb{Q}}, \mathbb{P}_2\otimes U_{\mathcal{R}}\otimes U_{\ell}\big)\\ &\quad\leqslant \Delta\big(F^*\hat{\mathbb{Q}}, \hat{\mathbb{Q}}_2\otimes U_{\mathcal{R}}\otimes U_{\ell}\big) + \Delta(\hat{\mathbb{Q}}_2\otimes U_{\mathcal{R}}\otimes U_{\ell}, \mathbb{P}_2\otimes U_{\mathcal{R}}\otimes U_{\ell}) &(7)\\ &\quad\leqslant (\varepsilon-\rho) + \Delta(\hat{\mathbb{Q}}_2, \mathbb{P}_2) &(8)\\ &\quad\leqslant \varepsilon - \frac{1}{2}\rho. &(9)\end{aligned}$$

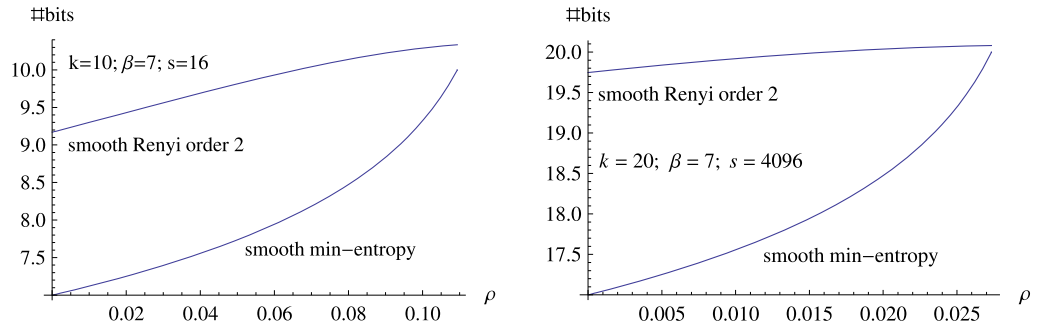

Fig. 1. Comparison of $\mathsf{H}_\infty^{\rho,\mathrm{strict}}(X)$ and $\mathsf{H}_2^{\rho,\mathrm{strict}}(X)$ in a toy example: $X \in \{0,1\}^k$, s elements have probability $p_1 = 2^{-k}(1+\beta)$, and the remaining $2^k - s$ elements all have probability $p_2 = (1 - sp_1)/(2^k - s)$.

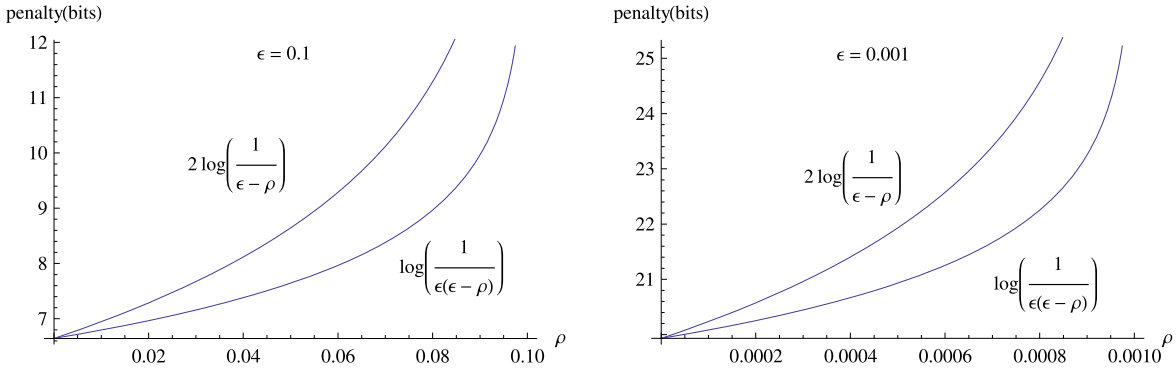

Fig. 2. Comparison of the penalty terms in the unconditional case.

In (8) we used the uniformity property of F. In (9) we applied Lemma 7. It follows that

$$\Delta\left(F^*\mathbb{P}, \mathbb{P}_2 \otimes U_{\mathcal{R}} \otimes U_\ell\right) \leqslant \varepsilon. \qquad \square$$

Compared to the proof of Theorem 1 there is one extra ingredient: the extra triangle inequality required to handle the distance between the marginals $\mathbb{P}_2$ and $\hat{\mathbb{Q}}_2$.

3.1. Extractable randomness in the unconditional case

3.1.1. Comparison of Theorem 1 with prior results

- The smooth Rényi entropy of order 2 ($\mathsf{H}_2^{\rho,\mathrm{strict}}$) is larger than the smooth min-entropy ($\mathsf{H}_\infty^{\rho,\mathrm{strict}}$) that features in (1). Fig. 1 shows for a simple example distribution that the difference can be several bits.
- The smoothening 'penalty' paid in the logarithm term is less severe than the $-2\log(\varepsilon - \rho)$ penalty in (1), as is shown in Fig. 2. The difference is $-2\log(\varepsilon-\rho) + \log\varepsilon(\varepsilon-\rho) = -\log(1-\rho/\varepsilon)$, which can amount to several bits in practice.

For low-entropy sources, the relative improvement due to these two differences can be substantial.

3.1.2. Effect of different smoothening definitions

What changes when another version of smooth entropy is considered? The two smoothening-related parts of the proof are the use of the un-normalized version of the leftover hash lemma (Lemma 4), and the final step (4), where the statistical distance between $\hat{\mathbb{Q}}$ and $\mathbb{P}$ enters. Let us first consider the loosely bounded $\mathsf{H}_2^{\rho,\mathrm{loose}}(\mathbb{P})$. Here, just as in the 'strict' case, the norm of $\hat{\mathbb{Q}}$ is $1-\rho$, and $\Delta(\mathbb{P},\hat{\mathbb{Q}})$ is $\frac{1}{2}\rho$. Hence, in terms of $\mathsf{H}^{\rho,\mathrm{loose}}$ we get exactly the same statements as in Theorem 1.

Next we consider the 'normalized' $\mathsf{H}_2^{\rho,\mathrm{norm}}$. Now we have to use $\alpha = 0$ in Lemma 4 since $\hat{\mathbb{Q}}$ is normalized. Furthermore we have $\Delta(\mathbb{P},\hat{\mathbb{Q}}) \leqslant \rho$ from Definition 11 as opposed to $\frac{1}{2}\rho$ in the other two cases. Following all the steps of the above proof yields the following result.

$$\ell_{\mathrm{ext}}^{\varepsilon}(X) \geqslant \max_{\rho\in[0,\varepsilon)}\left[\mathsf{H}_2^{\rho,\mathrm{norm}}(\mathbb{P}) + 2 - 2\log\frac{1}{\varepsilon-\rho}\right]. \tag{10}$$

How does this compare to Theorem 1? For most[5] distributions $\mathbb{P}$, the smooth entropies $\mathsf{H}_2^{\rho,\mathrm{norm}}(\mathbb{P})$ and $\mathsf{H}_2^{\rho,\mathrm{strict}}(\mathbb{P})$ will be roughly the same, since both definitions of smoothening allow one to remove an amount ρ from the highest probabilities $\mathbb{P}(x)$. The logarithmic penalty term of (10), however, is more severe: $\log\frac{1}{(\varepsilon-\rho)^2} \geqslant \log\frac{1}{\varepsilon(\varepsilon-\rho)}$. Hence Theorem 1 gives a sharper bound.

3.2. Extractable randomness in the conditional case

3.2.1. Comparison of Theorem 2 with prior results

- The expression $\widetilde{\mathsf{H}}_{2(1|2)}^{\rho,\mathrm{strict}}(\mathbb{P})$ is larger than $\mathsf{H}_{\infty(1|2)}^{\rho,\mathrm{strict}}(\mathbb{P})$ (see Lemma 8) for two reasons. (a) Rényi entropy of order 2 is larger than min-entropy (see Fig. 1); (b) taking the average over y is more favorable than taking the worst-case y.
- The factor $\sqrt{1-\rho}$ gives a slight improvement over (1) by reducing the penalty term. In fact, this factor was already present in an expression for the distance between mixed quantum states in [14, Corollary 5.6.1 on p. 88], but the consequences for extractable randomness were never explicitly stated. In practice ρ is so small that $\sqrt{1-\rho}\approx 1$, i.e. the improvement is negligible.

3.2.2. Effect of different smoothening definitions

We look what changes occur in the proof of Theorem 2 when the 'strict' smoothening is replaced by another type. In the 'loose' version nothing changes, since $\hat{\mathbb{Q}}$ will be the same (i.e. an amount ρ removed from the highest peaks), and $\Delta(\mathbb{P}_2,\hat{\mathbb{Q}}_2)$ as well as $\Delta(\mathbb{P},\hat{\mathbb{Q}})$ remain unchanged.

From the 'loose' version we directly obtain a bound on the 'normalized' version: Since the vicinity B_{norm}^{ρ} is a subset of $B_{\mathrm{loose}}^{2\rho}$ we can write $\mathsf{H}^{2\rho,\mathrm{loose}} \geqslant \mathsf{H}^{\rho,\mathrm{norm}}$. This yields the following result.

Theorem 3. *Let $\mathbb{P}$ be a probability measure on $\mathcal{X}\times\mathcal{Y}$, and $(X,Y)\sim\mathbb{P}$. Let $\varepsilon>0$. Then*

$$\ell_{\mathrm{ext}}^{\varepsilon}(X|Y) \geqslant \max_{\rho\in[0,\varepsilon/2)}\left[\widetilde{\mathsf{H}}_{2(1|2)}^{\rho,\mathrm{norm}}(\mathbb{P}) + 2 - \log\frac{1-2\rho}{(\varepsilon-2\rho)^2}\right].$$

How does this compare to Theorem 2? As in the unconditional case, for most distributions $\mathbb{P}$ it will hold that $\widetilde{\mathsf{H}}_{2(1|2)}^{\rho,\mathrm{norm}}(\mathbb{P}) \approx \widetilde{\mathsf{H}}_{2(1|2)}^{\rho,\mathrm{strict}}(\mathbb{P})$. The penalty term is larger. Hence Theorem 2 gives a sharper bound.

It must be remarked, however, that Eq. (10) and Theorem 3 can be useful in contexts where normalization must be strictly adhered to. For instance, it regularly occurs that a distribution $\mathbb{Q}$ is not exactly known, but is guaranteed to be at most ρ-distant from some distribution $\mathbb{P}$ with known properties. Bridging the gap from $\mathbb{Q}$ to $\mathbb{P}$ can be done by 'smoothening', but only if the normalized version of Definition 11 is applied.

4. Reduction of storage requirements by using almost universal instead of universal hash functions

An important practical consideration for implementing a randomness extractor is the available nonvolatile memory in a device. As mentioned in Section 1.1, the information reconciliation step requires storing the helper data. Furthermore, the privacy amplification step also needs storage space. This may not be immediately obvious from the previous sections. In this step the device needs an implementation of an appropriate hash function. The proofs of Theorems 1 and 2 require the use of a *universal* family of hash functions mapping an input $x\in\mathcal{X}$ to a smaller target space $\mathcal{T}$. As can be seen in Definition 18 (p. 1189), these hash functions are labeled by a random variable $R\in\mathcal{R}$; the value of R needs to be stored. Both the helper data and the R value are 'unique' in the sense that they are different for each instance of a source and for each enrolment. Hence these pieces of data are preferably stored in flash or eeprom memory. In resource-constrained devices this type of memory is expensive. Typical constructions of a universal family of hash functions require that $\log|\mathcal{R}|$ is (almost) as large as $\log|\mathcal{X}|$. For instance, the construction of Example 8.39 in [17] requires

$$\#\mathrm{bits} = \log|\mathcal{R}| = \log|\mathcal{X}| - \log|\mathcal{T}|. \tag{11}$$

For highly non-uniform sources X this is prohibitive.

It is possible to save on nonvolatile memory by relaxing the constraints on the hash function: By allowing small deviations from perfect universality, it becomes possible to have a tradeoff between the quality of the privacy amplification and the space needed to store R. Definition 19 introduces *η-almost universal* families of hash functions.[6] There are constructions [17] of such hash functions, with $\eta=(1+\delta)/|\mathcal{T}|$, requiring only

$$\log|\mathcal{R}| = \mathcal{O}\left(\log|\mathcal{T}| + \log\frac{\log|\mathcal{X}|}{\log|\mathcal{T}|} + \log\frac{1}{\delta}\right). \tag{12}$$

[5] As long as $\mathbb{P}$ is not too close to uniform.

[6] Exact universality occurs at $\eta=1/|\mathcal{T}|$.

Comparing (12) to (11) we see that the dependence on $|\mathcal{X}|$ has changed from $\log|\mathcal{X}|$ to $\log\log|\mathcal{X}|$, which is much smaller. The incurred penalty $\log(1/\delta)$ is manageable.

Example. Consider a case where $\mathcal{X} = \{0,1\}^{1024}$ and $\mathcal{T} = \{0,1\}^{128}$. For universality we need $\log|\mathcal{R}| = 896$ bits of storage, while for almost universality with $\delta = 2^{-16}$ we only need approximately 147 bits.

Finally we mention that the reduced quality of the privacy amplification due to a finite δ does not lead to significant worsening of the extractable randomness (see [20]).

Theorem 4. *Let $\varepsilon \geqslant 0$. Let $X \sim \mathbb{P}$ be a random variable on $\mathcal{X}$. Let $\{\Phi_r\}_{r\in\mathcal{R}}$ be an η-almost universal family of hash functions from $\mathcal{X}$ to $\{0,1\}^\ell$, with $\eta = (1+\delta)2^{-\ell}$. Then the amount of randomness that can be extracted from X using this family of hash functions is bounded from below by*

$$\ell^\varepsilon_{\mathrm{ext}}(X) \geqslant \max_{\rho\in[0,\varepsilon-\delta/[4\varepsilon])} \left[\mathsf{H}_2^{\rho,\mathrm{strict}}(\mathbb{P}) + 2 - \log\frac{1}{\varepsilon(\varepsilon-\rho)-\delta/4} \right].$$

Proof. Same as for Theorem 1, but without setting $\delta = 0$. □

Theorem 5. *Let $\varepsilon \geqslant 0$. Let $(X,Y) \sim \mathbb{P}$ be random variables on $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let $\{\Phi_r\}_{r\in\mathcal{R}}$ be an η-almost universal family of hash functions from $\mathcal{X}$ to $\{0,1\}^\ell$, with $\eta = (1+\delta)2^{-\ell}$. Then the amount of randomness that can be extracted from X, conditioned on Y, using this family of hash functions is bounded from below by*

$$\ell^\varepsilon_{\mathrm{ext}}(X|Y) \geqslant \max_{\rho\in[0,\varepsilon-\sqrt{\delta}/2)} \left[\widetilde{\mathsf{H}}_{2(1|2)}^{\rho,\mathrm{strict}}(\mathbb{P}) + 2 - \log\frac{1}{(\varepsilon-\rho)^2-\delta/4} \right].$$

Proof. Same as for Theorem 2, but without setting $\delta = 0$. □

The effect of nonzero δ on the extractable randomness is negligible as long as $\delta \ll \varepsilon^2$.

5. Summary

We have improved the known lower bound on the extractable entropy from non-uniform sources in the unconditional case (Theorem 1) and in the conditional case (Theorem 2). In the unconditional case, the improvement stems from using H_2 instead of H_∞ and from using an un-normalized version of the leftover hash lemma. In the unconditional case, the improvement comes from using the *smooth average conditional Rényi entropy* $\widetilde{\mathsf{H}}_{2(1|2)}^{\rho,\mathrm{strict}}$, which is the best match to the entities appearing in the leftover hash lemma.

We have investigated three of the definitions of smoothening with respect to their impact on the extractable entropy bounds. It turns out that the two un-normalized versions are equivalent in this respect, and yield a sharper bound than the normalized version.

We have studied the amount of nonvolatile memory that is needed to store the random label (R) of the hash functions employed in privacy amplification. A significant reduction is known to be achieved ($\log|\mathcal{X}|$ to $\log\log|\mathcal{X}|$) by switching from a universal family of hash functions to an η-almost universal family. We have shown that the penalty paid in the extractable randomness can be made vanishingly small.

Acknowledgment

We thank one of the anonymous reviewers for a slight sharpening of Theorem 3.

Appendix A

Proof of Lemma 4. This proof follows the approach in [17], but for an un-normalized distribution $\mathbb{Q}$. We use Lemma 3, replacing $\mathbb{Q}$ in the lemma by $f^*\mathbb{Q}$ and $\mathcal{Z}$ by $\mathcal{R}\times\{0,1\}^\ell$. We have to compute the collision probability c,

$$c = \sum_{r\in\mathcal{R}}\sum_{t\in\{0,1\}^\ell} \left[\left(f^*\mathbb{Q}\right)(r,t)\right]^2 = \sum_{r\in\mathcal{R}}\sum_{t\in\{0,1\}^\ell} \left[\frac{1}{|\mathcal{R}|}\sum_{x\in\mathcal{X}:\ f(r,x)=t} \mathbb{Q}(x)\right]^2.$$

We introduce the notation $\langle A\rangle$, for some predicate A, to denote 1 if A is true and 0 if A is false. This allows us to write

$$c = |\mathcal{R}|^{-2} \sum_{x,x' \in \mathcal{X}} \mathbb{Q}(x)\mathbb{Q}(x') \times \sum_{r \in \mathcal{R}} \sum_{t \in \{0,1\}^\ell} \langle f(r,x) = t \rangle \langle f(r,x') = t \rangle$$
$$= \frac{1}{|\mathcal{R}|} \sum_{x,x' \in \mathcal{X}} \mathbb{Q}(x)\mathbb{Q}(x') \Pr\big[f(R,x) = f(R,x')\big].$$

In the last expression the probability is with respect to the random variable R, while x and x' are fixed. Next we split the x, x' summation into a part where $x' = x$ and a part where $x' \neq x$, and we apply the defining property of the almost universal hash function f.

$$c = \frac{1}{|\mathcal{R}|}\Bigg(\sum_{x \in \mathcal{X}} \mathbb{Q}^2(x) + \sum_{x,x' \in \mathcal{X}:\, x' \neq x} \mathbb{Q}(x)\mathbb{Q}(x') \Pr\big[f(R,x) = f(R,x')\big] \Bigg)$$
$$\leqslant \frac{1}{|\mathcal{R}|}\Bigg(\sum_{x \in \mathcal{X}} \mathbb{Q}^2(x) + \sum_{x,x' \in \mathcal{X}:\, x' \neq x} \mathbb{Q}(x)\mathbb{Q}(x') \frac{1+\delta}{2^\ell} \Bigg)$$
$$= \frac{1}{|\mathcal{R} \times \{0,1\}^\ell|}\Bigg(2^\ell \sum_{x \in \mathcal{X}} \mathbb{Q}^2(x) + (1+\delta)\Bigg[(1-\alpha)^2 - \sum_{x \in \mathcal{X}} \mathbb{Q}^2(x) \Bigg] \Bigg)$$
$$\leqslant \frac{1}{|\mathcal{R} \times \{0,1\}^\ell|}\Bigg(2^\ell \sum_{x \in \mathcal{X}} \mathbb{Q}^2(x) + (1-\alpha)^2 + \delta \Bigg).$$

Substitution into Lemma 3 yields Lemma 4. □

Proof of Lemma 5. We begin by introducing the notation $\mathbb{Q}_{1|y} \in \mathcal{P}_\mathcal{X}$, where $\mathbb{Q}_{1|y}(x) := \mathbb{Q}_{1|2}(x,y)$ and $\sum_{x \in \mathcal{X}} \mathbb{Q}_{1|y}(x) = 1$ for all y. The statistical distance in the lemma can be expressed as

$$\Delta\big(F^*\mathbb{Q}, \mathbb{Q}_2 \otimes U_\mathcal{R} \otimes U_\ell\big) = \sum_{y \in \mathcal{Y}} \mathbb{Q}_2(y) \Delta\big(F^*\mathbb{Q}_{1|y}, U_\mathcal{R} \otimes U_\ell\big). \tag{13}$$

We apply Lemma 4 (with $\alpha = 0$, since $\mathbb{Q}_{1|y}$ is normalized) to get

$$\Delta\big(F^*\mathbb{Q}_{1|y}, U_\mathcal{R} \otimes U_\ell\big) \leqslant \frac{1}{2}\sqrt{\delta + 2^{\ell - \mathsf{H}_2(\mathbb{Q}_{1|y})}}. \tag{14}$$

We substitute this into (13) and then apply Jensen's inequality for concave functions (Lemma 1). This yields

$$\Delta\big(F^*\mathbb{Q}, \mathbb{Q}_2 \otimes U_\mathcal{R} \otimes U_\ell\big) \leqslant \frac{1}{2}\sqrt{\sum_{y \in \mathcal{Y}} \mathbb{Q}_2(y)} \sqrt{\sum_{y \in \mathcal{Y}} \mathbb{Q}_2(y)\big[\delta + 2^{\ell - \mathsf{H}_2(\mathbb{Q}_{1|y})}\big]}$$
$$= \frac{1}{2}\sqrt{1-\beta}\sqrt{\delta(1-\beta) + 2^{\ell - \widetilde{\mathsf{H}}_{2(1|2)}(\mathbb{Q})}}.$$

Lemma 5 follows. □

References

[1] B. Barak, R. Shaltiel, A. Wigderson, Computational analogues of entropy, in: APPROX 2003 + RANDOM 2003, in: Lecture Notes in Comput. Sci., vol. 2764, Springer-Verlag, Berlin, 2003, pp. 200–215.

[2] J.D.R. Buchanan, R.P. Cowburn, A. Jausovec, D. Petit, P. Seem, G. Xiong, D. Atkinson, K. Fenton, D.A. Allwood, M.T. Bryan, Forgery: 'fingerprinting' documents and packaging, Nat. Brief Commun. 436 (July 2005) 475.

[3] L. Carter, M.N. Wegman, Universal classes of hash functions, J. Comput. System Sci. 18 (2) (1979) 143–154.

[4] D. Clarke, B. Gassend, M. van Dijk, S. Devadas, Secure hardware processors using silicon physical one-way functions, in: R. Sandu (Ed.), ACM CCS '02, 2002.

[5] G. DeJean, D. Kirovski, Radio frequency certificates of authenticity, in: IEEE Antenna and Propagation Symposium – URSI, 2006.

[6] Y. Dodis, R. Ostrovsky, L. Reyzin, A. Smith, Fuzzy extractors: How to generate strong keys from biometrics and other noisy data, Cryptology ePrint Archive, Report 2003/235, 2003.

[7] Y. Dodis, R. Ostrovsky, L. Reyzin, A. Smith, Fuzzy extractors: How to generate strong keys from biometrics and other noisy data, SIAM J. Comput. 38 (1) (2008) 97–139.

[8] J. Guajardo, S.S. Kumar, G.J. Schrijen, P. Tuyls, FPGA intrinsic PUFs and their use for IP protection, in: P. Paillier, I. Verbauwhede (Eds.), CHES, in: Lecture Notes in Comput. Sci., vol. 4727, Springer, 2007, pp. 63–80.

[9] J. Hastad, R. Impagliazzo, L.A. Levin, M. Luby, A pseudorandom generator from any one-way function, SIAM J. Comput. 28 (4) (1999) 1364–1396.

[10] T. Holenstein, R. Renner, One-way secret-key agreement and applications to circuit polarization and immunization of public-key encryption, in: V. Shoup (Ed.), Advances in Cryptology – CRYPTO '05, in: Lecture Notes in Comput. Sci., Springer-Verlag, August 2005, pp. 478–493.

[11] T. Holenstein, R. Renner, On the randomness of independent experiments, available at http://arxiv.org/abs/cs.IT/0608007, August 2006.

[12] J.-P.M.G. Linnartz, P. Tuyls, New shielding functions to enhance privacy and prevent misuse of biometric templates, in: J. Kittler, M. Nixon (Eds.), Conference on Audio and Video Based Person Authentication, in: Lecture Notes in Comput. Sci., vol. 2688, Springer-Verlag, 2003, pp. 238–250.

[13] R. Pappu, B. Recht, J. Taylor, N. Gershenfeld, Physical one-way functions, Science 297 (September 2002) 2026–2030.
[14] R. Renner, Security of quantum key distribution, PhD thesis, Swiss Federal Institute of Technology (ETH) Zurich, September 2005, available at http://arxiv.org/abs/quant-ph/0512258.
[15] R. Renner, S. Wolf, Simple and tight bounds for information reconciliation and privacy amplification, in: B. Roy (Ed.), Advances in Cryptology – ASIACRYPT 2005, in: Lecture Notes in Comput. Sci., vol. 3788, Springer-Verlag, December 2005, pp. 199–216.
[16] R. Renner, S. Wolf, J. Wullschleger, The single-serving channel capacity, in: Proceedings of the International Symposium on Information Theory (ISIT), in: IEEE, July 2006, available at http://arxiv.org/abs/cs.IT/0608018.
[17] V. Shoup, A Computational Introduction to Number Theory and Algebra, Cambridge University Press, 2008, version 2, available at http://www.shoup.net/ntb/ntb-v2.pdf.
[18] D.R. Stinson, Universal hashing and authentication codes, in: J. Feigenbaum (Ed.), CRYPTO, in: Lecture Notes in Comput. Sci., vol. 576, Springer, 1991, pp. 74–85.
[19] D.R. Stinson, Universal hashing and authentication codes, Des. Codes Cryptogr. 4 (4) (1994) 369–380.
[20] D.R. Stinson, Universal hash families and the leftover hash lemma, and applications to cryptography and computing, J. Combin. Math. Combin. Comput. 42 (2002) 3–31.
[21] P. Tuyls, G.J. Schrijen, B. Škorić, J. van Geloven, R. Verhaegh, R. Wolters, Read-proof hardware from protective coatings, in: L. Goubin, M. Matsui (Eds.), Cryptographic Hardware and Embedded Systems – CHES 2006, in: Lecture Notes in Comput. Sci., vol. 4249, Springer-Verlag, 2006, pp. 369–383.
[22] P. Tuyls, B. Škorić, T. Kevenaar, Security with Noisy Data: Private Biometrics, Secure Key Storage and Anti-Counterfeiting, Springer, 2007.
[23] B. Škorić, T. Bel, A.H.M. Blom, B.R. de Jong, H. Kretschman, A.J.M. Nellissen, Randomized resonators as uniquely identifiable anti-counterfeiting tags, March 2008, Secure Component and System Identification (SECSI) Workshop.